The development of genetic genealogy tools adapted to an endogamous context: its usefulness for research on migration, by James Smith, president, association Atelier Généalogique, Marseillle, France

In my genetic genealogy research from 2016 onwards, I encountered the problem of endogamy when searching for the birth parents of a genealogy pupil and friend, Francine: in short, endogamy [1] produces a surfeit of cousins and namesakes, blurring the lines.

Six years of intensive research – literally thousands of hours of work - have resulted in the creation of a large family tree centered on one location, Corato (as well as on some surrounding towns), in Puglia, southern Italy. In 2022, there are almost 24,000 people in it. The family tree is backed up by a database of more than 20,000 photos of records, stored and consultable on Google Drive. These numbers have been reached through building out family trees for over 100 people, mostly Francine's DNA matches. Of these, more than 60 were Americans of Italian descent. Hundreds of researchers on various genealogical sites have copied from our tree, the *albero genealogico coratino*[2] (on the websites of Ancestry, MyHeritage, Geneanet, and Filae).

However, our project has a historical, as well as a genealogical focus. No in-depth family history research about people from Corato can ignore the migratory phenomenon, and the genealogical research conducted by the descendants of emigrants is a lot richer if it includes the country where the people came from. Emigrants from Corato started going to the US from 1902, but France took over as the main destination from about 1920.

Francine and I translated and had published in a Franco-Italian version a book by the historian Pasquale Tandoi entitled 'Quando i clandestini eravamo noi'. This is still available on demand, and an English translation, *When we were the illegal immigrants*, is available on our web-site[3]. Information about our genealogical research spread and engendered a project entitled 'A tale of two twinned cities: Corato and Grenoble' (with apologies to Charles Dickens !). Essentially, we aim to improve access to resources that would be useful for both genealogical research and migration studies about Corato ; and to generally promote exchanges between descendants of Coratini emigrants, encourage the writing of testimonies, facilitate access to archives *in loco.*

The main focus of our project is to promote the writing of the history of *coratino* emigration to the area around Grenoble. A sizeable proportion of the population of Corato[4] opted to settle in France, mostly in Grenoble, and it is often said that the *coratini* contributed much to the industrial, socio-economic and cultural development of this major French city. At the same time, we are mindful of the fact that members of the same families that settled in the Grenoble area emigrated to various other destinations (USA, Venezuela, Brazil, Germany, etc), thereby no doubt providing many examples of the complex phenomenon of 'transnational families'[5].

Today our project is monitored by a scientific committee under the presidency of Professor Biagio Salvemini, full professor of Modern History at the Aldo Moro University in Bari. An example of our work was presented to the Oxford University Migrations Programme in May 2022[6] .

---

[1] A calculation of the level of endogamy in Corato, based on the study of marriages during the 1835-65 period, carried out by the Atelier Généalogique in collaboration with the VISUGED association, showed that Corato was 97% endogamous.
[2] https://www.ancestry.co.uk/family-tree/tree/167791321/recent?usePUBJs=true
[3] https://www.emigrazione-corato.org/medias/files/when-we-were-the-illegal-immigrants2.pdf
[4] Tandoi found that 7,450 people out of a total of 13066 emigrants who applied for passports made their way to France between 1920 to 1959…and he did not count illegal immigrants to France. In 1921, the population of Corato was 50 010 (cf Tandoi) ; in 2022, it is 48,253.
[5] Cf *Famiglie transnazionali*, in P.Corti, M.Sanfilippo, a cura di, *Storia d'Italia. Annali 24, Migrazioni*, Torino, Einaudi,2009, pp.303-316.
[6] https://www.emigrazione-corato.org/medias/files/summaries-submitted-for-publication-on-the-compas-blog.pdf

An experiment aimed at better identifying the relevant cousins in an endogamous context

Wikitree [7], an American genealogical website, is presently engaged in the development of an application called 'X-friends' (developer : Greg Clarke) that may well be a significant contribution, not only to genetic genealogical research but also to migration studies based on the 'memory' of migration, in other words, on family history. This bi-disciplinary approach aimed at collecting the testimonies of people who emigrated would be topical from a historiographic point of view. More than 100 years after the main migratory flux, many taboos about telling the story of migration, and its multi-faceted nature, are being lifted.

The genealogical experiment we are conducting in cooperation with Wikitree concerns the continued search for Francine's birth parents by adapting genetic genealogy research methods to take into account the endogamous context of Corato. Our hypothesis is that the methodology, if it proves to be successful, could be relevant to the search for any person or persons in an endogamous context.

In principle, since the transmission pattern of the X-chromosome is very specific (father to daughters only, mother to all her children) the tracking of potential carriers of the X – both 'upwards' to several 'terminal ancestors' and 'downwards' to all the people who may have inherited the X from the latter – will reduce the number of relevant cousins to be taken into consideration during the search, by eliminating certain genealogical male lines.

Potential X-matches can be usefully traced for anyone in a family tree if the latter is large enough and is accurately built. Accuracy is, of course, essential : if the family tree is flawed, the potential X-matches will be imaginary. Wikitree's strict sourcing standards would normally prevent this, and a DNA confirmation tool developed jointly by Wikitree and Gedmatch, as well as other genetic genealogy tools, should help to prevent errors.

The technical procedure of the X-friends experiment can be described as follows:

1. From the *albero genealogico coratino*, for each of the main autosomal DNA matches (80cM[8] or more) of the subject (i.e. Francine), the Family Tree Maker software exports a Gedcom[9] of his/her direct ancestors and their descendants for 9 generations.

2. With its X-friends application[10], Wikitree extracts from each of the Gedcoms a .csv file containing the names of potential X matches identified by a single continuous bottom-up, then top-down operation of the "widget" designed to track the transmission paths of these X chromosomes. This .csv file is created starting from a person having taken a DNA test selected from the family tree produced by the Gedcom.

3. Merging several of these .csv files should identify potential carriers of the same X-chromosomes as Francine from several sources, and thus indicate which pedigree lines are most promising.

The above procedure was repeated 16 times, producing 16 .csv files. The pedigree chart below illustrates how relevant cousins can be identified by finding people who potentially carry an X-match

---

[7] Wikitree is an American society described as a *« free, shared social-networking genealogy website that allows users individually to research and to contribute to their own personal family trees while building and collaborating on a singular worldwide family tree within the same system »* (Wikipedia).

[8] cM: centimorgans, a unit of measurement of DNA shared with another person.

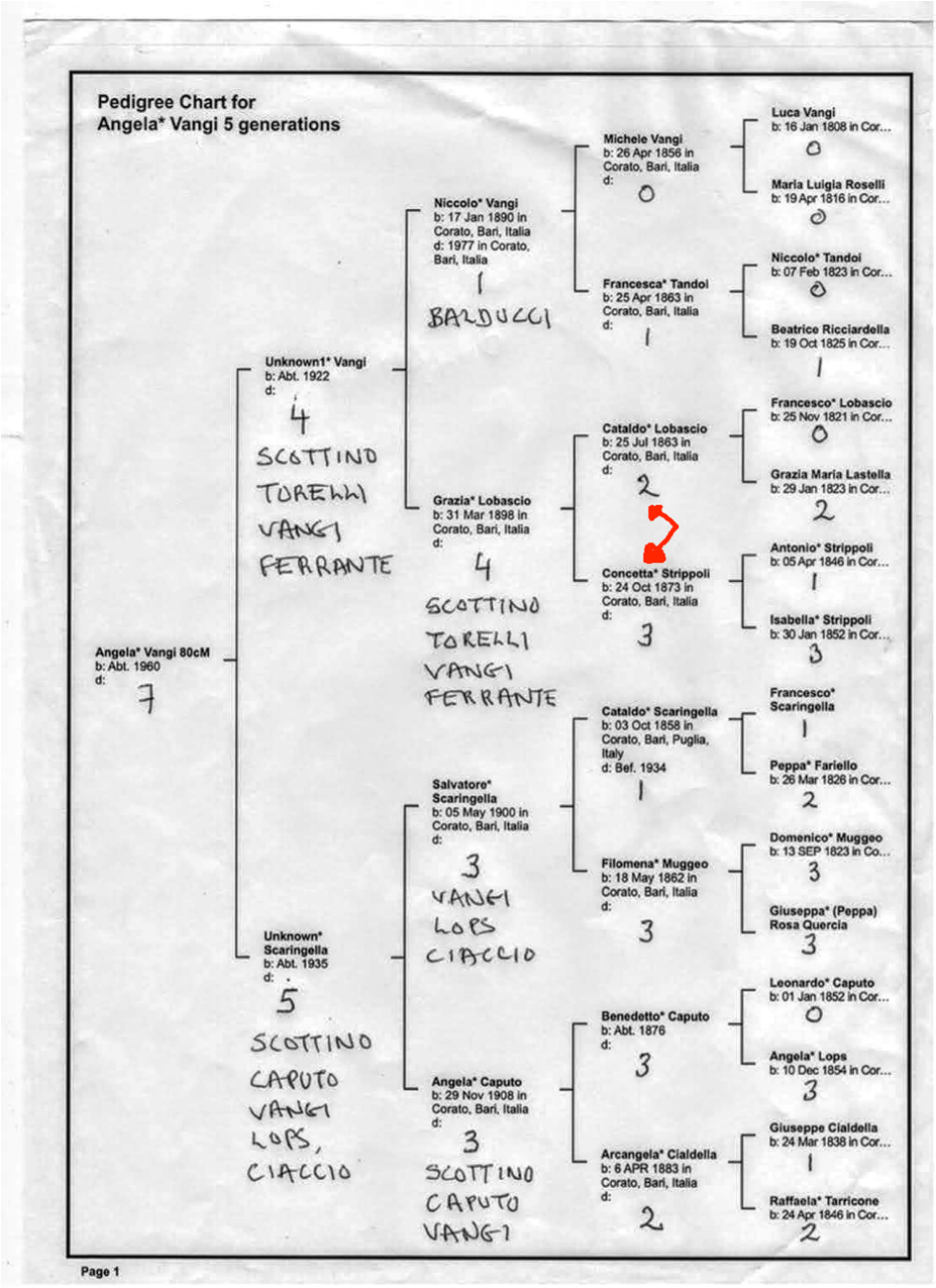
[9] GEDCOM files are sequential files in text format, which allow the exchange of computerized data between different genealogy software. A Gedcom is created from a family tree and enables the transfer of this tree from one genealogy software to another. Gedcoms can also be made from an appropriate .csv file (see below).

[10] https://apps.wikitree.com/apps/clarke11007/Xfriends.php

with Francine from different sources. Angela Vangi, for example, an 80cM autosomal match with Francine, potentially shares X-chromosomes with Francine from 7 different sources.

However, the X-friends app is able to identify all potential X-Match carriers among the descendants of any person's ancestors, not just people who have taken a DNA test. The chart below indicates, for example, that Cataldo Lobascio and Concetta Strippoli are potential carriers of the X from 2 and 3 sources respectively. By building out the descendancy of this particular couple, it is still possible to identify other people who were (or are) potential carriers of the X from several sources, but who have not taken a DNA test.

Our hypothesis is that by converting the resulting .csv files to Gedcoms and thereby creating a specific family tree in which the people who have taken a DNA test are included, a pattern may emerge indicating the most promising genealogical lines of Francine's direct ancestors. Our hypothesis is also that the X-friends app could help to verify the identity of emigrants in a similar way.



Relevance of Wikitree's X-friends app to our project's overall strategy aimed at improving access to resources useful for both genealogy and migration studies

The experiment with Wikitree is on-going ; in other words, the process described above is not yet complete. The conversion of .csv files of potential carriers of X-matches to gedcoms[11], and the merging of the family trees thus produced, will result in a specially focused genealogical family tree. If the experiment is successful, our *albero genealogico coratino* could be used by anyone researching their family history in Corato or destination countries. The same methodology could be applied with any large tree based on a one-place study of an endogamous location. It could become part of a bi-disciplinary approach to the history of emigration, linking genealogical and historical perspectives. Our hypothesis is therefore that genealogy can find its place in a concept of history based on 'memory', by identifying migrants and their descendants capable and willing to testify.

In the Italian context, the identification process is made difficult not only by endogamy but by lack of access to parish records and restrictions imposed by the *Stato Civile* – where birth records are freely available only up to 1910 – as well as by legislation on what Italians call *privacy* (some of which is increasingly imposed by the EU). *Stato Civile* records also have many imprecisions and cases of mistaken identity, aggravated by illiteracy. The ability to identify people during the 1910-1930 period is especially important because it is a time when many migrants were born.

Fundamentally, if the history of migration is to be written on the basis of the 'memory' of it, in other words on family histories, genealogical research must find ways of overcoming these restrictions. Of course, there are other ways of accessing genealogical information : ship manifests, expired work and residence permits, censuses, university memoirs and theses, newspaper articles, naturalization and military data, family archives, etc. As a means of verifying identity, Wikitree's 'X-friends' app could well be a very useful addition to existing genetic genealogy tools.

Marseille, October 2022

[11] The Atelier Généalogique is working on this with Jean Chabaud, the designer of the VISUGED software (https://www.visuged.org/index.htm) which was used in March 2022 to calculate the level of endogamy in Corato.